

ARM Data Quality Office

K. Kehoe, A. Theisen, J. Monroe, J. King, R. Peppler

E. Wong, M. Schneider, N. Myers, J. Quinn, T. Kloss, E. Green, A. Duling, R. Bunker, J. Anthony



ARM Data Quality Process

Instrument Deployment and Maintenance



- Mentors work with site to install and calibrate instrument. Document all actions in logs.
- Site personnel perform daily rounds checking instrument function in the field and on instrument computer display. Document all actions in logs.
- Mentors or site personnel enter Data Quality Problem Report at any point in this process when needed.



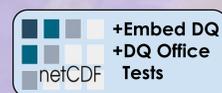
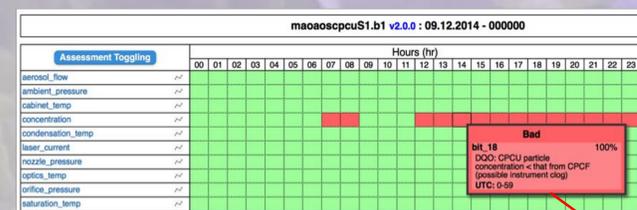
Data Processing



Data Management Facility processes raw data into community standard files and higher order Value Added Products (VAP).

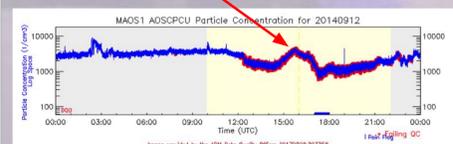
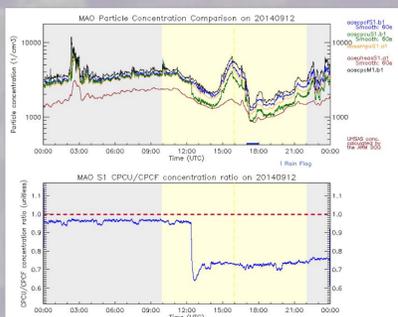
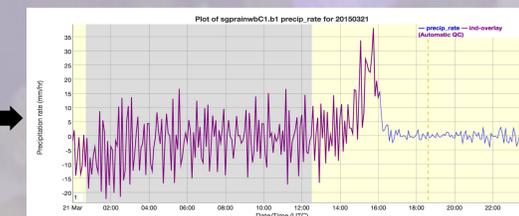
- Apply calibration
- Add embedded quality control fields including advanced quality control to Value Added Products
- Monitor data flow
- Operators enter Data Quality Problem Report at any point in this process when needed.

Data Inspection



Embedded Quality Control Field User Guide

- Introduction
- Examples in Various Languages
 - Examples using IDL
 - Examples using R
 - Examples using Python 2
 - Example using Python 2 with NumPy
 - Examples using Python 3
 - Examples using PERL
 - Examples using Matlab



Data Quality Problem Report

ROUTINE Data Quality Problem Report (DQPR): 4315

Issue Date: 9/18/2014
 DQPR Originator: IM/OPS - Allison Aiken
 Date Closed: 11/19/2014
 Locations: MAO - S1: Manacapuru, Amazonas, Brazil; MAOS
 Instrument Class: CPC

Time Range of Data Quality or Data Problem:

Start Date: (MM/DD/YYYY) 09/12/2014 Time: 12:00 UTC
 End Date: (MM/DD/YYYY) Time: UTC

Submit above DQ Time changes

Current QA Code Selected: Questionable Data Update QA Code

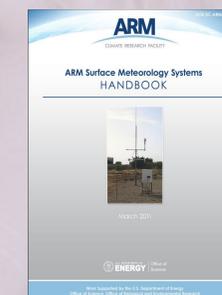
- DQ Office metrics summarize embedded quality control and additional DQ Office tests
- The DQ Office metrics tables are designed for web-browsing speeds using hourly quality flag summaries and pre-generated plots
- DQ Analysts review data daily to weekly creating weekly Data Quality Assessment reports
- Analysts and other ARM personnel use interactive web-tools to zoom into a problem when needed
- Fully automated processing, and e-mail alerts
- On-site processing for remote deployments - transfer smaller sized metrics and plots files only
- DQ Analysts enter Data Quality Problem Report at any point in this process when needed.

Data Quality Office in the Numbers: 2016

- Total Plots and Metrics Created or Modified
 - Plots: 7,076,620
 - Metrics: 533,995
- 996 Data Quality Reports submitted (versus 222 submitted in 2011)
- DQ Analysts submitted 8,378 Data Quality Assessments (versus 2,799 in 2011)
- Submitted 482 Data Quality Problem Reports
 - 342 have currently resulted in Data Quality Reports

Data User Resources

Instrument Handbooks Provide Summary and Detail



- Instrument Handbooks and Value Added Product technical reports provide information including operation, maintenance, data descriptions and data quality.
- Example excerpt from AERIPROF VAP:
 - Below is a description of quality control checks for both the input data for the AERIPROF retrieval algorithm and the retrieved AERI profiles.
 - Input File Quality Control for Lidar data:
 - 0: Lidar data OK
 - 1: Fog/condensation on lidar window
 - 2: No data within AERI specified time period
 - 3: No valid data found within LIDAR file

Embedded Quality Control Fields Mask Incorrect or Suspect Data: User's Option

Data users copy-paste code into their code to enable filtering incorrect or suspect data using embedded quality control.

Data Quality Reports Mask Incorrect or Suspect Data: User's Option

Data Quality Report

DQR ID: D141017.4
 DQR SUBMITTER: Chongqi Kuang
 DQR SUBJECT: MAO/CPFC/S1 - CPCU under-counting
 COMMENTS: One approach to correcting for this lower sample flow rate in the CPCU is to identify sampling periods where it is likely that no sub 10 um aerosol are being sampled. During those time periods, the CPFC and CPCU measured number concentrations should, under normal operating conditions, be the same, and the CPFC measurements can then be used to calibrate the CPCU sampling flow rate. Over time, the decrease in CPCU sample flow rate can be tracked and used to correct the CPCU reported number concentration.

QUALITY COLOR: Red - Incorrect
DQR DESCRIPTION:
 Under normal operating conditions, the CPCU measured number concentration should be equal to or greater than the CPFC measured number concentration. The fact that the CPCU is under-counting with respect to the CPFC during this time period strongly suggests that there is a clog in the sampling capillary, resulting in a lower than expected sample flow rate and therefore a lower than expected measured number concentration.

DQR DATASTREAM(S):
 maooscpus1.a1
 concentration

START DATE/TIME: 01/24/2014 11:30:00 UTC END DATE/TIME: 02/03/2014 01:00:00 UTC
 START DATE/TIME: 02/04/2014 17:00:00 UTC END DATE/TIME: 03/25/2014 18:00:00 UTC

Option 2:

- Users copy-paste code into their code to filter data with Data Quality Report web-service.
- Web-service does most of the work for user's code.

DQR Web Service User Guide

- Introduction
- Examples in Various Languages
 - Examples Using R
 - Examples Using Python 2
 - Examples Using Python 3
 - Examples Using IDL
 - Examples Using PERL
 - Examples Using Matlab
- Appendices
 - Parameter Descriptions
 - Status Codes
 - DQR Details



Option 1:

- Mask data within data file using Data Quality Reports. User selectable options in Data Discovery Tool.
- Repeatable if outcome is not what user needs.

